Designing for Ease of Collaborative Effort in Human-Robot Joint Action

Kerstin Fischer, Lars C. Jensen, Franziska Kirstein

Institute for Design and Communication
University of Southern Denmark
Sonderborg, Denmark
kerstin@sdu.dk

ABSTRACT

Joint action between humans and robots can be improved either by transferring more and more human skills into the robot, decreasing the difference between these unequal partners, or by designing human-robot interactions in such a way that people can adjust to these collaborations, making use of their natural tendency to orient at the ease of collaborative effort, i.e. taking over those tasks that are very difficult for robots but easy for humans. In this paper, we explore what conditions need to be satisfied so that humans make use of their intuitive capabilities in human-robot interaction. The case studies presented concern robot learning from demonstration and illustrate a) that users have intuitive knowledge and possibly even methods that they bring to bear seamlessly and b) that probing into such intuitive knowledge can be difficult if people do not understand that their behavior matters.

Keywords

joint action, learning from demonstration, collaboration, asymmetric communication

1. INTRODUCTION

Joint action does not necessarily mean that all collaborators have the same capabilities and carry out the same tasks; humans and robots have, for instance, very different strengths and weaknesses.

One possibility to improve human and robot joint action is to provide robots with more and more human capabilities in order to bridge the gap between the unequal partners. Previous work shows that providing robots with human capabilities, such as pro-active multi-state perspective taking (Pandey et al. 2013) or contingent response (e.g. Fischer et al. 2013), indeed makes interactions increasingly smooth and natural.

Another possibility to bridge the gap in asymmetric interactions such as between humans and robots is to make use of a natural human capability, namely to attend to ease of collaborative effort. That is, when people are collaborating with each other, they seamlessly take over certain tasks if they are difficult for the communication partner. For instance, in a series of experiments, Schober (1998, 2009) shows that people take their partners' perspective, even though this means more cognitive effort for them, if they notice that their partners have more problems than they are having themselves. Furthermore, in speech to infants and young children, caregivers take over almost all interactional work (Filipi 2009). Clark & Wilkes-Gibbs (1986) therefore introduce the principle of ease of collaborative effort to account for the finding that people generally orient less at minimizing their own effort but instead rather strive to minimize the joint effort.

In our first case study, users subconsciously produce clues to a distinction that is easy for humans and difficult for robots, namely whether the user is demonstrating an action for the robot or is working on his or her own, solitary task. Since the robot is meant to learn to collaborate from the observation of the human, knowing whether it currently needs to observe the human action or not is crucial, and thus such clues can be very helpful. However, as our second case study shows, getting people to employ their intuitive knowledge may not be trivial at all. If people are not aware that there is a problem for the robot or when they don't understand that their behavior matters, interfaces for the human-robot collaboration need to be designed that allow people to make use of their intuitive knowledge in ways that the robot can process.

The case studies we present here both concern learning from demonstration, also known as programming by demonstration (e.g. Billard et al. 2008; Dillmann et al. 1995).

2. STUDY I: INTUITIVE DISPLAYS

In this study (cf. also Kirstein et al. 2015), we describe a situation in which human users switch between collaborative work and solitary tasks. Only the former involve human-robot interaction and, in this case, also robot learning from demonstration. In order to collaborate effectively, it is crucial that the robot distinguishes between these two types of phases in the workflow. What we have found was that users intuitively employ cues to the status of their current actions. For the robot, making such a distinction would be very difficult. However, naïve users seamlessly take over this collaborative work by using a reliable to cue to the collaborative phases: they smile. Thus, this study is an example for successful joint action where human users orient at ease of collaborative effort and simply take over the task.

The data were elicited on a collaborative assembly task during which naïve human users were asked to instruct a robot to hand over the appropriate parts for the assembly of a wooden toolbox, which the user then had to assemble him- or herself. The experiments were carried out using the Third-Hand-Platform at the Institute of Computer Science at the University of Innsbruck within the frame of the 3rd Hand Project (Lioutikov et al 2014).

The robotic platform used comprises a Kuka robot arm equipped with a Schunk hand and a head with two inbuilt cameras and one Kinect camera on top. Other cameras were placed around the robot's workspace to record the sessions for later analysis and for the wizard to control the robot. The workspace of the robot is a table in front of the robot with a foam layer on top. For this experiment, the six pieces of the wooden box were placed on the table in a holding system so that the robot could easily grasp them. The participants stood at the other side of the table, opposite the robot and were equipped with the tools (drill, screws and instruction manual) to assemble the box. Behind the participants, an engineer ensured everyone's safety. Another engineer, not visible to the participants, controlled the robot from behind a screen.

30 participants (aged 18-38) were included in the analysis, 8 female and 22 male. Participants were recruited by word-of-mouth and rewarded with a bar of chocolate for their time.

The task was to guide the robot to assist them in assembling a wooden toolbox. All participants received the same introduction and had to find out themselves how they could instruct the robot to fetch the parts. After introducing the participants to the task, the facilitators did not intervene except when assisting users with the drill. The robot reacted based on the users' gestures (not, for instance, their speech). The participants' interactions with the robot were video recorded. After participants completed the task, they filled out a questionnaire about their interaction with the robot.

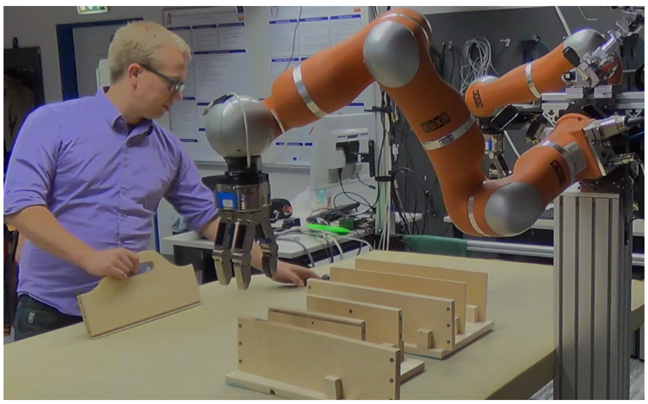


Figure 1. The experimental set-up

We noticed that participants smiled whenever they turned to the robot. We thus analyzed the recorded sessions systematically concerning the participants' smiling behavior when they interacted with the robot compared to when they assembled the wooden box themselves. Interactions of the participants with the facilitators were disregarded from the analysis. For every situation, the number and length of participants' smiles was counted. The participants' interaction with the robot was further divided into two kinds of situations: situations in which participants instructed the robot to fetch the parts and situations in which the robot handed the parts over to the participants. Since the assembly involved six parts, there are six instruction and six handover situations per participant. The participants' behavior before and after they smiled was analyzed to determine the contexts in which smiling occurred.

The analyses reveal that participants smile significantly more when interacting with the robot than when assembling the toolbox. All in all, participants smiled around five times more when interacting with the robot (236 times) than when assembling the wooden box (42 times). The quantitative analysis shows that in both phases in which participants interact with the robot, instructing and handovers, participants smile more often and for a longer time than during assembly. T-testing reveals that the solitary assembly phases differ significantly from both other conditions such that people smile more when interacting with the robot (for details, see Kirstein et al. 2015).

To sum up, in this study, people had intuitive knowledge about the status of the interaction that was not readily available to the robot. However, people systematically employ social signals to mark off the two different phases, providing the robot with reliable cues to the current interactional status.

3. STUDY II: DESIGNING FOR COLLABORATION

In our second study (Fischer et al. in preparation), people are also taken to possess intuitive knowledge, but here they do not make it available easily to the robot. The problem addressed here is the choice of the robot control point during programming by demonstration. A very efficient way to demonstrate actions to a robot is teleoperation (Chernova and Thomaz 2014; Campbell et al. 1996; Dillmann et al. 1995). Teleoperation is a *"demonstration technique in which the teacher operates the robot learner platform and the robot's sensors record the execution"* (Argall et al. 2009: 473). When actions are demonstrated to a robot by means of teleoperation, the robot is controlled by means of a reference point that is coordinated with the teleoperation device. Usually this point is the tool center point (TCP). All joints are operated in relation to that control point. During programming by demonstration, the robot control point is furthermore the relevant, since most informative, point the system records and learns from. The problem now concerns the fact that a fixed robot control point such as the TCP may not always be the most informative point, nor the best point for robot control. For instance, depending on whether the robot is moving only itself or itself and an object, the most informative control point will differ. While in the first case the control point of an industrial robot arm, for instance, is in the robot's 'hand', the control point should be somewhere in the object when the robot is carrying a somewhat lengthy object, such as a peg from the Cranfield benchmark set for peg-in-hole tasks (Collins et al. 1984).

In order to decide automatically on the appropriate control point, the robot would need to know whether it is holding an object or not, what the exact dimensions of the object are, how it is holding the object and what the current activity is – especially whether a particular part of the object is in focus, such as the lower part of a peg that is being inserted into a hole, or whether the robot is moving the whole object from one place to another.

Such decisions require considerable reasoning capabilities, which common industrial robot arms are not equipped with. In contrast, our hypothesis is that humans find this decision very easy to make, and that taking over this task in the interest of easing the collaborative effort would come with no particular costs for the human, yet greatly improves the interaction if the human user understands how he or she can contribute to the collaborative effort. We thus hypothesized that if users have the choice, they will intuitively inform the robot correctly about what control point to choose. The challenge then consists in providing users with the opportunity to make use of their intuitive knowledge.

The robot used for this study is the MARVIN (Multi Armed Robotic and Vision Intelligence) platform (Savarimuthu et al. 2013), located in a laboratory at the University of Southern Denmark. The platform comprises two Universal Robot UR5 arms with 6 DoFs, of which only one was used in the current experiments. Each arm is equipped with a three-finger gripper SDH-2. A numpad allows users to select a grasp mode for the robot gripper. A 6-axis force-torque sensor is mounted between the robot hand and the gripper.

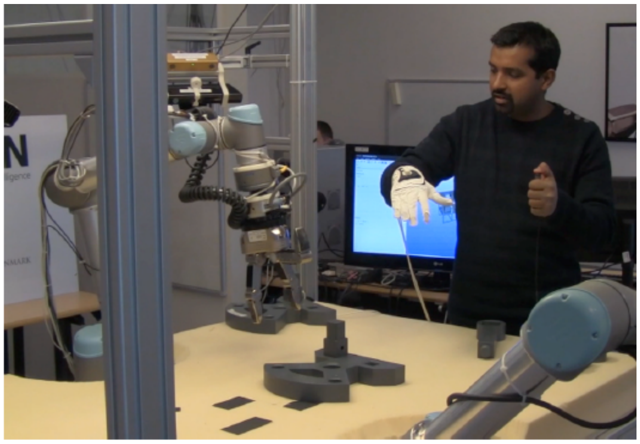


Figure 2: The Marvin platform

The experiments were recorded with two video cameras directed at both robot and human user, both of which were external to the platform. The participants in this experiment were recruited by word-of-mouth and ads in the cafeteria of our university.

3.1 Prestudy

We first carried out a study using the common data glove for teleoperating the robot (see Figure 2). In this pre-study, 16 participants aged 19-36, one of whom is female (average age 25.3), carried out a typical peg-in-hole task using a benchmark setting, the Cranfield set (Collins et al. 1984). Since we wanted to keep the tasks short in order to avoid both fatigue and a learning effect, users received only four pieces of the Cranfield set: a faceplate, a square peg, a round peg and a separator (see Figure 2). Their task was to insert the two pegs into the faceplate and to place the separator on top. On a scale from 1 to 4, participants selfreport a mean experience of 2.4 with robots and 2.8 concerning how often they play video games.

Participants using the glove need 4:49-10:53 minutes of actual moving time of the robot to complete the task, with an average of 188.24 seconds per peg (sd 121.45). Within the time frame of the 20 minutes given, only 10 of the 16 participants finish the task. Three only succeed to insert the two pegs, two manage to insert only one peg, and one participant does not even succeed with a single peg.

We asked most participants which part of the robot they believed they were controlling, and it turned out that they did not have a realistic idea of the parts they were controlling (see Figure 3). However, many of the participants in fact wished for different control points.

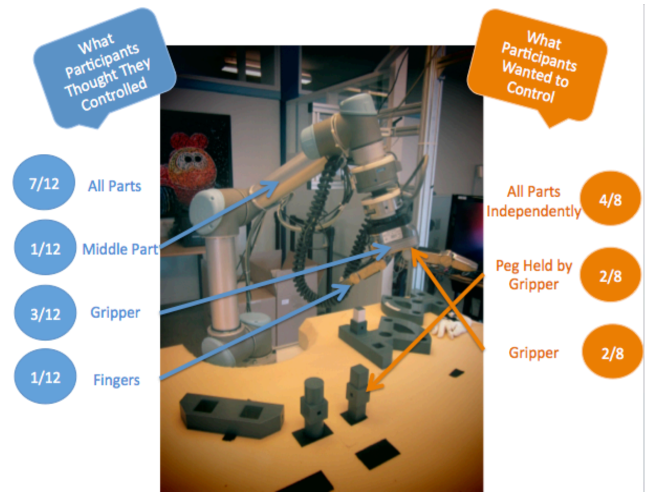


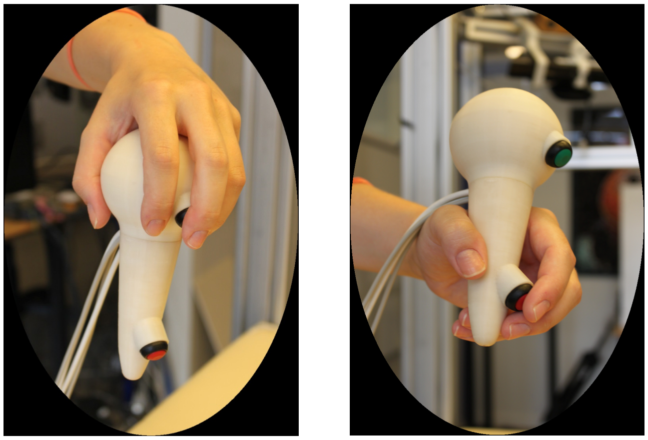
Figure 3: What participants thought they controlled and what they would have liked to control in the Data Glove Condition

These results show that people did not develop an accurate understanding of how the robot really works (the robot actually uses the tool center point above the gripper as the control point). They did however bring in an accurate view of what would be necessary, namely at least two different control points – as is also apparent from the fact that participants who use kinesthetic guidance, i.e. who manipulate the robot arm directly, would never use only a single control point but would move their hands to different places on the robot to control it. Thus, the problem, that during teleoperation there is usually only one robot control point, which is located in the tool center, did not become clear to the users.

To provide users with the possibility to switch between control points when they deemed necessary, we set out to develop a novel interface that allows users to do exactly what they wished to do, namely to switch between robot control points, to match their intuitions about which part to control.

3.2 Device Design

In order to make use of users' intuitive knowledge, we developed a novel teleoperation device that is designed to suggest different holds and thereby communicate to the robot which control point to choose. By intuitively choosing an appropriate hold for a given task, users are expected to choose intuitively the right control point as well.



Figures 4ab: The novel device

For the novel device (see Figures 4ab), we integrated two buttons that communicate two different control points to the robot. The two buttons correspond to two different reference points in the robot (see Figure 5). The device was shaped to invite two different ways of holding the device, which in turn invite two different kinds of activities:

a) moving the robot arm in large movements in a hold that resembles the shape of the robot arm (and of the way the data glove is held)

b) operating the robot to carry out finetuning, for instance, when positioning the respective object correctly in a hole. This hold resembles holding a pen (pens are held in this way to allow us to make fine movements with them).

Thus, the two holds are taken to be iconic of the kinds of activities to be carried out and thus to suggest implicitly which button to press (and correspondingly which control point to choose).

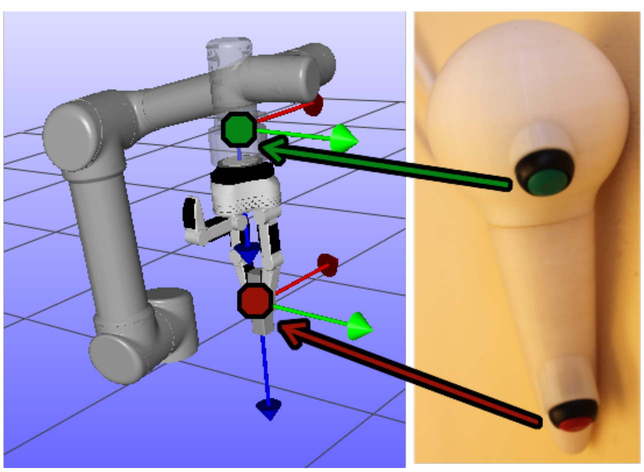


Figure 5: The two buttons on the novel teleoperation device correspond to two different robot control points

3.3 Pilot Study

The pilot study during which participants use the new teleoperation device comprises 16 participants aged 22-59 years, four of whom are female (average age 33.5). They selfreport an average experience with robots of 1.9 and 1.8 concerning the frequency with which they play video games.

Like in the pre-study, users were greeted, asked to fill out a consent form and shown an instruction video created just for this experiment, in which the task, the robot, and the handling of the respective teleoperation device and the numpad were explained.

In the video, concerning the new device, the two intended ways of holding the device were briefly demonstrated without further explanation. Then, users were brought into the lab space and introduced to the robot.

The videos were analyzed for efficiency, success and error rate for each participant in the two conditions. Efficiency was measured by calculating the time the robot was being moved, the success rate was determined by analyzing how many pegs and separators were correctly positioned within the 20-minute timeframe provided.

Regarding efficiency, it turns out that 15 of 16 participants complete the task, and the one participant who does not complete the whole task at least manages to insert both pegs. The range of time needed is 4:06 to 11:04 minutes, with an average of 142.22 seconds per peg (sd 82.25). Participants thus did not perform significantly faster than in the pre-study using the data glove, yet they were more successful.

However, it turned out that only 13% of the users made use of both control points, and that the other users rather chose one way to hold the device and then did not change any more during the experiments. Thus, even though teleoperation by means of the new device produced higher success rates, participants in fact did not make the use of the two buttons that correspond to the different reference points. Figure 6 shows that only 13% of the participants used both buttons and thus switched between reference points. Most of the other participants decided for reference point 1 (the grasp for large movements).

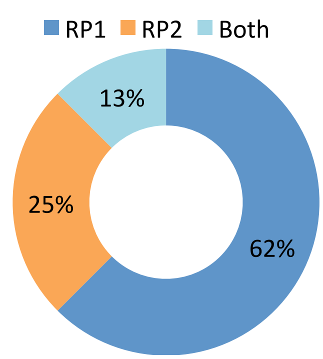


Figure 6: Percentage of participants using the upper button (RP1), the lower button (RP2) or both

Furthermore, when asked which point they believed they were controlling, participants responded in very similar ways as they did in the condition with the data glove as teleoperation device (see Figure 7).

Like when using the data glove, participants wished to be able to switch control points; however, they did not understand that the new device allowed them to do this. When analyzing the reasons for the failure of our device, we reviewed the introduction video, and we realized that when we introduced the two buttons that indicated the different control points to the robot, we had used the formulation 'you can operate the robot like this *or* you can operate it like that'; i.e. we had unintentionally presented the two ways of holding the device as equal alternatives. We therefore implicitly signaled to our participants that their choice of button did not matter and had no relevance. That users did not understand that the way they held the device had any impact is indicated by Figure 8, which shows some snapshots of ways in which participants held the device. Consequently, the introductory video may have been more misleading than helpful, and so we carried out a second set of experiments in which we presented users with a new introduction video that made the significance of the two buttons clear.

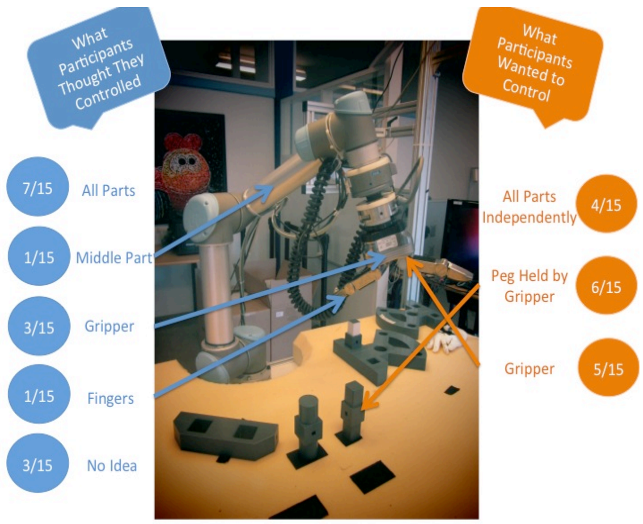


Figure 7: What participants thought they were controlling and what they would have liked to control in the New Device Condition

The results of this pilot clearly show that just providing users with the possibility to switch control points was not sufficient to make them use their intuitive knowledge for the robot. This may be due to lack of willingness to do so, or simply with the fact that they did not know that it was important – which made them foreground other considerations (such as level of comfort). In order to distinguish whether users' limit of orientation of ease of collaborative effort was reached and they simply did not care about whether the robot understood the correct control point or whether they did not understand from the introductory video that their choice mattered, we carried out another, final study.

3.4 Final Study

This study replicates the previous experiments, just with a new introductory video. This video introduced the function of the two buttons. We compare here the effects of using the data glove and the new device directly.

30 participants took part in this study, 15 in each condition. In condition 1, the data glove condition, there were 5 women and 10 men between 21 and 57 years of age (average age 27.9). On a scale from 1 to 4, participants selfreport an average experience with robots of 2.3 and with computer games 2.2. In condition 2, the condition in which participants used the new device, there were 2 women and 13 men between 18 and 39 years of age (average age 24.1). Participants selfreport an average experience with robots of 1.9 and with computer games 2.9.

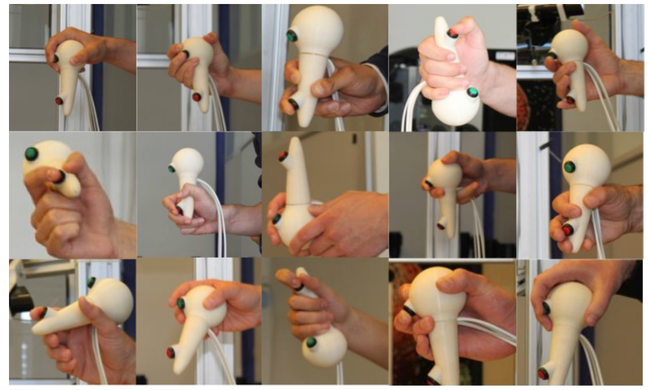


Figure 8: Different ways of holding the device in the pilot study

The analysis of the data shows that this time participants had understood the use of the two buttons and used the device in the way expected. Of 90 large movements that participants carried out in Condition 2 with the new device, all were done using reference point 1. Likewise, 74 times users employed the lower button to carry out fine movements. Thus, all participants switched between the buttons, which suggests that the video instruction was successful in this study.

The analysis of the quantitative data shows that demonstrations by means of the data glove take significantly longer, both in terms of the time in which the robot was moved and the total demonstration time, for the round peg and the separator than the demonstrations by means of the new device. Results are near significant for the square peg. Thus, using the novel device in the intended ways indeed contributed to the efficiency of the demonstrations. Consequently, when participants understand that they can choose the robot control point themselves, they all take the opportunity to do so. In this study, the video tutorial was sufficiently clear about the function of the novel device, and all users made use of both buttons. However, the different steps we present here indicate how difficult it was to make use of users' intuitions about the use of the device. Nevertheless, as soon as people understand that there is a problem, they all take over the extra interactional effort to improve the joint action.

4. CONCLUSION

We have presented two examples of human-robot joint action, in which users possess intuitive knowledge that can improve the collaboration, and like in asymmetric interactions among humans, users did take over extra interactional work in order to make the interaction more successful. However, the two case studies differ considerably: Whereas in the first study, users subconsciously produced concomitant signals to the robot as a byproduct of their understandings of the status of the current activity, in the second study, people possessed the same kind of knowledge about the type of activity at stake, namely large movement versus finetuning, yet since participants did not understand how the robot was controlled, they did not understand that there was a problem and hence that their choices mattered. Thus, users' understandings of the world, while useful as a resource for the joint collaboration, as the two studies show, can also prevent users' from making the right choices; in study II, since people did not understand the use of the handholds demonstrated to them and the problem that they were designed to solve, they rather went with their own explanations and motivations and ignored the recommendations given in the video in the pilot study. Thus, probing users' intuitive knowledge sometimes has to address such knowledge explicitly in order for participants to make use of it in a given collaboration. However, on the whole, users proved more than willing to take over extra tasks for the sake of ease of collaborative effort.

5. DESIGN IMPLICATIONS

The current discussion has shown that one way to improve human-robot joint action, also of the kind which serves as a common starting point for the current workshop, is to guide users into making use of their enormous experience in social cooperation and into taking over those activities that are particularly hard for robots to achieve. As we have seen, identifying possible contributions by humans can be achieved by observing closely what people do anyway, but it may also require sophisticated interface design to make use of human intuitive knowledge. A crucial step in the employment of such knowledge turned out to be the communication of the problems robots face as well as their exact functioning; consequently, much work in the design of joint action should go into the communication of the respective robot's affordances, especially its strengths and weaknesses, mechanisms and functionalities.

Another design implication of the current finding, that users attend to ease of collaborative effort also in human-robot interactions, is that not only are people willing to take over extra efforts to make the collaboration successful, but they may also expect robots to step back where their performance is far below the level of the human user. That is, the principle of ease of collaborative effort demands that the interaction partners collaborate on making the joint action most effective – which suggests that there may be problems if robots try to solve tasks that humans are particularly good at. Thus, it may make sense to determine beforehand what aspects of a joint action humans would prefer the robot to take over.

6. ACKNOWLEDGMENTS

We are grateful to Dadhichi Shuklar, Özgür Erkent and Justus Piater for their collaboration on Study I and Kamil Kuklinski, Thiusius Rajeeth Savarimuthu, Dorte Sølvason and Norbert Krüger for their help during the elicitation of the user studies reported on as Study II. Furthermore, we would like to thank Richard Beck for creating the prototype of the new device. This research was partially funded by the EU projects 3rd Hand and IntellAct, as well as by the DSF project Carmen.

7. REFERENCES

[1] Argall, B. Chernova, S., Veloso, M. and Browning, B. 2009. A survey of robot learning from demonstration*, Robotics and Autonomous Systems,* vol. 57, pp. 469-483.

[2] Billard, G., Calinon, S., Guenter, F. 2006. Discriminative and adaptive imitation in uni-manual and bi-manual tasks. *Robotics and Autonomous Systems, 54*, 370-84.

[3] Billard, A., Calinon, S., Dillmann, R. and S. Schaal, 2008. Robot programming by demonstration, *Springer Handbook of Robotics*. pp. 1371-1394.

[4] Calinon, S. 2009. A Probabilistic Approach, *Robotic Programming by Demonstration,* Lausanne: EPFL Press, 2009.

[5] Campbell, C.L., Peters II, R.A., Bodenheimer, R.E., Bluethmann, W.J., Huber, E. and R. O. Ambrose. 2006. Superpositioning of Behaviors Learned through Tele-operation, *IEEE Transactions on Robotics* 22, 1, 79-91.

[6] Chernova, S. and Thomaz, A. 2014. *Robot Learning from Human Teachers*. Morgan & Claypool Publishers.

[7] Clark, H.H. and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22 , 1–39.

[8] Collins, K., Palmer A. and K. Rathmill, 1984. The Development of a European Benchmark for the Comparison of Assembly, *Proc. First Robotics Conference*, Brussels.

[9] Demiris, Y. and G. Hayes. 2002. Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model, *Imitation in Animals and Artifacts,* Cambridge, MA: MIT Press, pp. 327-361.

[10] Dillmann, R. 2004. Teaching and Learning of Robot Tasks via Observation of Human Performance. *Robotic Autonomous Systems* 47, 2-3: 117-127.

[11] Dillmann, R. Kaiser, M. and Ude, A. 1995. Aquisition of elementary robot skills from human demonstration. *Int. Symposium on Intelligent Robotic Systems*, Oisa, Italy.

[12] Filipi, A. 2009. *Toddler and Parent Interaction*. Amsterdam: John Benjamins.

[13] Fischer, K, Katrin S.L, Saunders, J, Nehaniv, C, Wrede, B. and Rohlfing, K. 2013. The impact of the contingency of robot feedback on HRI. *Cooperative Technological Systems,* San Diego, May 20-24 2013.

[14] Goodwin, K. 2009. *Understanding Potential Users and Costumers*. In: *Designing for the digital age: how to create human-centered products and services*. Wiley.

[15] Kirstein, F., Fischer, K., Erkent, Ö. and Piater, J. 2015. Human Smile Distinguishes between Collaborative and Solitary Tasks in Human-Robot Interaction. *Late Breaking Results*, *Human-Robot Interaction Conference 2015*, Portland, Oregon.

[16] Kukliński, K., Sølvason, D., Fischer, K, Marhenke, I., Kirstein, F., aus der Wieschen, M., Krüger, N. and Savarimuthu, T.R. 2014. Teleoperation for Learning by Demonstration: Data Glove versus Object Manipulation for Intuitive Robot Control. *ICUMT*, St. Petersburg, Russia.

[17] Lioutikov, R., Kroemer, O., Peters, J. and Maeda, G. 2014. "Towards a third hand," in 1st International Workshop on Intelligent Robot Assistants.

[18] Pandey, A.K., Ali, M. and Alami, R. 2013. Towards Task-Aware Proactive Sociable Robot based on Multi-Stae Perspective Taking. *Intern. Journal of Social Robotics* 5, 215-236.

[19] Pardowitz, M., Knoop, S., Dillmann, R. and R. D. Zöllner, 2007. Incremental learning of tasks from user demonstrations, past experiences and vocal comments," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 37 (2)* pp. 322-332.

[20] Savarimuthu T., Liljekrans, D., Ellekilde, L-P, Ude, A.A., Nemec, B. and Krüger, N. 2013. Analysis of human peg-in-hole executions in a robotic embodiment using uncertain grasps, In Proceedings of the *9th International Workshop on Robot Motion and Control, RoMoCo*, 2013.

[21] Schober, M.F. 1998. Different kinds of conversational perspective-taking. In S. C. Fussell and R. Kreuz (Eds.), Social and Cognitive Psychological Approaches to Interpersonal Communication , pp. 145–174. Hillsdale: Lawrence Erlbaum.

[22] Schober, M.F. 2009. Spatial dialogue between partners with mismatched abilities. In K. Coventry, T. Tenbrink, and J. A. Bateman (Eds.), *Spatial Language and Dialogue*. Oxford: Oxford University Press.